Evaluating Teamwork in a Simulated Obstetric Environment

Pamela J. Morgan, M.D., C.C.F.P., F.R.C.P.C.,* Richard Pittini, M.D., M.Ed., F.R.C.S.C.,† Glenn Regehr, Ph.D.,‡ Carol Marrs, R.N.,§ Michèle F. Haley, B.A.‖

Background: The National Confidential Enquiry into Maternal Deaths identified "lack of communication and teamwork" as a leading cause of substandard obstetric care. The authors used high-fidelity simulation to present obstetric scenarios for team assessment.

Methods: Obstetric nurses, physicians, and resident physicians were repeatedly assigned to teams of five or six, each team managing one of four scenarios. Each person participated in two or three scenarios with differently constructed teams. Participants and nine external raters rated the teams' performances using a Human Factors Rating Scale (HFRS) and a Global Rating Scale (GRS). Interrater reliability was determined using intraclass correlations and the Cronbach α. Analyses of variance were used to determine the reliability of the two measures, and effects of both scenario and rater profession (R.N. *vs.* M.D.) on scores. Pearson product–moment correlations were used to compare external with self-generated assessments.

Results: The average of nine external rater scores showed good reliability for both HFRS and GRS; however, the intraclass correlation coefficients for a single rater was low. There was some effect of rater profession on self-generated HFRS but not on GRS. An analysis of profession-specific subscores on the HFRS revealed no interaction between profession of rater and profession being rated. There was low correlation between externally and self-generated team assessments.

Conclusions: This study does not support the use of the HFRS for assessment of obstetric teams. The GRS shows promise as a summative but not a formative assessment tool. It is necessary to develop a domain specific behavioral marking system for obstetric teams.

◆ This article is accompanied by an Editorial View. Please see: Murray DJ, Enarson C: Communication and teamwork: Essential to learn but difficult to measure. ANESTHESIOLOGY 2007; 106:895–6.

* Professor, ‖ Research Assistant, Department of Anesthesia, Women's College Hospital, University of Toronto. † Assistant Professor, Department of Obstetrics and Gynecology, § Educator, Birthing Unit, Department of Nursing, Sunnybrook Health Sciences Centre, University of Toronto, Canada. ‡ Professor, Wilson Centre for Research in Education, University of Toronto.

Address correspondence to Dr. Morgan: Department of Anesthesia, Women's College Hospital, 76 Grenville Street, Toronto, Ontario Canada M5S 1B2. pam.morgan@utoronto.ca. Individual article reprints may be purchased through the Journal Web site, www.anesthesiology.org.

Safety Attitudes Questionnaire. Available at: http://www.uth.tmc.edu/schools/med/imed/patient_safety/survey&tools.htm. Accessed November 20, 2006.

THE Institute of Medicine report entitled *To Err is Human* has fueled a compelling movement into patient safety initiatives and the mitigation of human error in medicine.[1] By acquiring an understanding of how medical teams perform, educational strategies can be developed to improve team performance and decrease the likelihood of errors.[2]

Evidence from safety research in high-risk organizations has demonstrated that nontechnical skills or behaviors must be studied because these cognitive and social skills have a pivotal role in maintaining safety, especially in critical care areas.[3–5] Evaluation of nontechnical skills is necessary for the assessment of both individual performances and group effectiveness, as well as to critically appraise the impact of training interventions. Equally important is the study of the evaluation tool and its ability to produce a valid and psychometrically robust measure of these performances.[6] Although behavioral marking systems have been developed for assessment of individual physician behaviors, there are no validated marking systems available for the assessment of obstetric teams.[7,8]

In aviation, safety attitudes of flight crews have been assessed by the Flight Management Attitudes' Questionnaire adapted for application to operating room teams, and referred to as the Operating Room Management Attitudes' Questionnaire (ORMAQ).[9–12] The ORMAQ has gone through many iterations and is now available for multiple user groups under the title Safety Attitudes Questionnaire.# Both the ORMAQ and the Safety Attitudes Questionnaire have been demonstrated to have acceptable internal consistency and have been used in research studies worldwide.[12–15] The ORMAQ was created to tap into the general teamwork, communication, stress recognition, and safety concerns of teams and has been used by investigators to develop behavioral tools to assess the individual performance of anesthesiologists and surgeons. It remains unclear to what extent these evaluation tools are domain specific and to what extent they apply to group as well as individual performance.[7,8] The purpose of this study was to determine whether an adaptation of the ORMAQ, titled the Human Factors Rating Scale (HFRS), and a Global Rating Scale (GRS) could be used to reliably assess obstetric team performance. The HFRS is a lengthy checklist consisting of 45

907

items used to measure such constructs as leadership, assertion, information sharing, and teamwork. The GRS uses a five-point scale with anchored descriptors to give an overall view of team performance.

Materials and Methods

Research ethics board approval from Sunnybrook Research Ethics Board, Toronto, Ontario, Canada, was received for this study.

Simulation Center

The simulation center was set up as an obstetric operating room and equipped with all necessary gowns, gloves, drapes, and instruments needed to perform a cesarean delivery. An external fetal heart rate monitor was present and provided a fetal heart tone if applied to the abdomen. Digital videography allowed as wide a view of the operating room as possible. Scenarios were videotaped using a video cassette recorder with vital signs superimposed on the image using a video mixer.

A patient chart was available and included prenatal records, nursing partograms, history, physical findings, laboratory data, and relevant consultative notes. Fetal heart rate traces since admission were available and reflected any abnormality that might have occurred.

Laerdal SimMan (Laerdal Medical Canada Limited, Toronto, Ontario, Canada) was used as the patient mannequin. To optimize the realism of the obstetric portion of the scenarios, an obstetric model was constructed that fit over the SimMan abdomen. The model resembled the size and shape of a term uterus, and its exterior was covered with material allowing the obstetricians to make either a vertical or a Pfannenstiel incision and to close the incision once finished. The abdomen could be prepped and draped, and once an incision was made, the fetus or fetuses and placenta could be delivered. The interior of the model had fenestrated tubing through which massive obstetric blood loss could be simulated. Depending on the scenario, blood loss could be minimized using the usual surgical techniques. A urinary catheter could also be inserted into the mannequin and urine output measured.

Scenarios

The National Confidential Enquiry into Maternal Deaths in the United Kingdom was consulted to identify the most frequent events leading to maternal demise.[16] In addition, obstetricians, anesthesiologists, and obstetric nurses from an academic institution were surveyed to determine what obstetric cases would be the most useful to rehearse using high-fidelity simulation. Answers were collated, data from the National Confidential Enquiry into Maternal Deaths were incorporated, and scenarios were developed using the most commonly cited emer-

gency situations. Scenarios commonly involved multiple events requiring management by the obstetric team. Four scenarios were developed and included (1) urgent cesarean delivery for parturient with worsening preeclampsia: critical event involving severe hypertension and pulmonary edema; (2) profound fetal bradycardia secondary to occult abruptio placenta: critical event involving massive blood loss and inability to maintain blood pressure; (3) emergency cesarean delivery in parturient with twin gestation at 34 weeks for umbilical cord prolapse: critical event amniotic fluid embolism; and (4) morbidly obese parturient with nonreassuring fetal heart rate trace; decision to perform cesarean delivery: difficult intubation, hypoxemia, and cardiac arrest. Further details of the scenarios can be found in appendix A.

Measurement Tools

Two primary measurement tools designed to assess team performance in obstetric emergency situations were evaluated in this study:

1. Human Factors Rating Scale: The HFRS is a behaviorally based performance evaluation scale minimally adapted for the obstetric context from the ORMAQ designed by Helmreich.[9,10,12] The HFRS contains 45 items related to five themes: leadership–structure, confidence–assertion, information sharing, teamwork, and error. Level of agreement with each statement is made using a five-point Likert scale where 1 = strongly disagree, 2 = slightly disagree, 3 = neutral, 4 = slightly agree, and 5 = strongly agree (appendix B).
2. Global Rating Scale of Performance: The GRS is a performance-based evaluation of overall team performance on the scenario that uses a single five-point rating scale with 1 representing an unacceptable performance, 3 representing an acceptable performance, and 5 representing a superior performance. Descriptive anchors are provided for each scale number and address the issue of error and patient safety (appendix C).

All scenario participants and raters completed a basic demographic questionnaire.

Scenario Participants

All staff obstetricians, anesthesiologists, and obstetric nurses at a single academic institution were invited to participate in the study. Subjects were enrolled on a first-come, first-serve basis and were given information packages about the nature of the study.

External Reviewers

Nine healthcare professionals were recruited to act as external video reviewers. The healthcare professionals were selected because of their expertise in the obstetric

environment or as human factors experts. The number of reviewers was predetermined on the basis of available funding to compensate video reviewers for their time.

Procedure

Three independent sessions were held, with each session involving two anesthesiologists, two obstetricians, five obstetric nurses, one obstetric resident, and one anesthesia resident. Each session involved the completion of four scenarios. For each scenario, a team of five or six members was constructed from the set of obstetric nurses, physicians, and residents in obstetrics and anesthesia present. The construct of the teams was achieved using computer-generated random number assignments. Across the four scenarios in a given session, each person participated in two or three scenarios with differently constructed teams.

On the day of the session and after informed consent was obtained, participants were given a 30-min orientation to the simulated operating room, surgical table and instruments, mannequin, drug cart, and anesthetic gas machine. After the orientation was completed, subjects were asked to complete the demographic questionnaire.

After the orientation, the four scenarios were run sequentially. For each scenario, participants were assigned to the team, and one of the nurses on the team was given a short synopsis of the patient problem by one of the investigators without the other members of the team present. The nurse then attended to the patient in the simulated operating room, where the patient's physical findings, including fetal heart rate, were simulated according to the scenario events. The action of the team from the initial introduction of the nurse to the patient was videotaped. While members of the first team independently completed the HFRS and GRS ratings of their own team's performance, the second team managed the second scenario. When the second scenario was completed, these team members completed the HFRS and GRS ratings of their team's performance. Teams were reassigned, and the two final scenarios were completed in the same way. Each scenario lasted approximately 20 min. Immediately after each scenario, participants were encouraged to self-reflect on both their own performance and their impact on the group. After all the scenarios were completed, one of the investigators reviewed one videotape of each team's performance with the participants. Standard crisis resource management techniques were used to debrief the sessions.

After completion of the three sessions, the 12 videotapes (three for each of the four scenarios) were sent to nine raters who independently evaluated the 12 videotaped team performances using the HFRS and the GRS.

Statistical Analysis

The interrater reliability of the HFRS and GRS when used by external raters viewing videotapes of the team performances was assessed for a single rater using intraclass correlation coefficients and for the average of all raters using the Cronbach α. In addition, the mean external ratings of performances on the four scenarios were compared using one-way repeated measures analysis of variance.

Similarly, the interrater reliability of the HFRS and GRS when generated as a self-assessment by team participants was assessed for the average of all six team members using the Cronbach α. Because each team involved different members, the mean self-generated ratings of performances on the four scenarios were compared using one-way between subjects analysis of variance. In addition for the self-generated ratings, analysis of variance was used to assess the effect of profession on ratings. Finally, self-assessed ratings for each team were compared to the externally generated scores using Pearson product–moment correlation coefficients.

Results

Thirty-four participants, 16 nurses, 6 obstetricians, 6 anesthesiologists, and 6 residents participated in 12 simulations, producing 71 self-generated HFRSs and GRSs for analysis. In total, these numbers represented more than 70% of the physicians involved in obstetric care at a single institution and a cross-section of registered nurses from the same environment with varying years of service, thereby reflecting the usual clinical teams who would normally respond to such events. Nine external raters completed the HFRS and GRS on each of the 12 videotaped performances.

Table 1 outlines the means and SDs of the team scores for each of the four scenarios using the HFRS and GRS when generated by the nine external raters.

Across the nine external raters evaluating the videotaped performances, the single-rater intraclass correlation coefficient for the HFRS was 0.341, suggesting that a single rater's scoring of the 12 performances was not highly predictive of any other individual rater's scoring. However, the nine-rater intraclass correlation coefficient (Cronbach α) for the HFRS was 0.823, suggesting that the average of nine raters was sufficient to generate a reasonably stable HFRS score for each team performance. The single rater intraclass correlation coefficient for the GRS across the nine raters evaluating the videotaped performances was slightly higher at 0.446, with a nine-rater Cronbach α of 0.879. The Pearson product–moment correlation between the HFRS and the GRS scores for the 12 scenarios (averaged across all raters) was 0.934, suggesting that the two measures were tapping largely the same construct in team performance. Analysis of variance for both the HFRS ($F_{3,24} = 8.09$, $P < 0.01$) and the GRS ($F_{3,24} = 16.89$, $P < 0.01$) revealed a significant difference in the ratings of performances

Table 1. Means (SDs) of the Team Scores

	HFRS External Raters	GRS External Raters	HFRS Self-assessed	GRS Self-assessed
Scenario 1	3.12 (0.35)	2.98 (0.64)	3.40 (0.51)	3.00 (0.97)
Scenario 2	3.53 (0.22)	4.07 (0.52)	3.42 (0.53)	3.61 (0.70)
Scenario 3	3.07 (0.51)	2.50 (0.97)	3.41 (0.31)	2.83 (0.86)
Scenario 4	3.51 (0.37)	3.52 (0.84)	3.54 (0.41)	3.40 (0.97)
	$F_{3,24} = 8.09$, $P < 0.01$	$F_{3,24} = 16.89$, $P < 0.01$	$F_{3,67} = 0.36$, NS	$F_{3,67} = 2.93$, $P < 0.05$

Scores for each of the four scenarios using the Human Factors Rating Scale (HFRS) and Global Rating Scale (GRS) when generated by nine external raters viewing videotapes and by participants self-rating immediately after the team performance.

NS = not significant.

among the four scenarios, suggesting that for some scenarios, it may be consistently more difficult for a given team to demonstrate effective team performance.

For the scenarios participants' self-ratings of team performance (generated independently by each of the six team members immediately after their team performances), the six-rater Cronbach α for team members' self-assessment scores was 0.15 for the HFRS and 0.74 for the GRS. This suggests that the average of six team members' ratings was moderately stable for the GRS but was problematically low for the HFRS. Similarly, analysis of variance revealed no significant difference between scenarios for HFRS scores ($F_{3,67} = 0.36$, not significant) but a borderline difference for GRS ratings ($F_{3,67} = 2.93$, $P < 0.05$), indicating the GRS, but not the HFRS, identified some scenarios as more difficult than others in a manner consistent with the external raters. Analysis of scores by profession indicated that nurses gave significantly higher team scores than physicians on the HFRS ($F = 6.26$, $P < 0.05$) but not on the GRS ($F = 2.96$, not significant). An analysis of profession-specific subscores on the HFRS revealed no interaction between profession of rater and profession being rated ($F = 1.98$, not significant), suggesting both groups rated the two professions' performances similarly. Finally, the Pearson correlation between self-generated scores and externally generated scores across the 12 performances was 0.24 for the HFRS and slightly higher at 0.44 for the GRS.

Demographic information is found in table 2.

Discussion

Behavioral marking systems have been used with simulation to provide formative and summative evaluation of individual behaviors that are otherwise difficult to assess.[3,17–19] Currently, however, there is no firm consensus on how to measure teamwork, with a lack of empirical data to validate measures.[20]

The transformation of an aviation tool to operating room teams (Flight Management Attitudes Questionnaire to the ORMAQ to the Safety Attitudes Questionnaire) has provided the means not only to address team safety attitudes, but to potentially be able to examine team performance using an adaptation of the tool.[9–12,15] Although the idea of adapting an existing safety attitudes tool to medical teams is an attractive one, this study has raised doubts about the efficacy of the tool in practice. Because the themes of the tool, such as leadership, confidence assertion, information sharing, and teamwork, seem on the surface to be important behavioral aspects of a team's performance, there may have been too many items within each category to allow for a

Table 2. Demographic Questionnaire

	Participants	Raters
Sex, %		
Female	70.6	88.9
Male	29.4	11.1
Age, mean ± SD, yr	40.89 ± 10.25	46.67 ± 10.55
Years in practice, mean ± SD	14.76 ± 9.63	16 ± 12.67
Previous educational experience using high-fidelity simulation, %		
Yes	23.5	33.3
No	76.5	66.6
Number of educational conferences attended in past 5 years, mean ± SD	5.66 ± 6.06	—
Use interactive Web-based learning for continuing education, %		
Yes	38.2	—
No	55.9	—
Previous use of virtual reality simulation for procedural skills training, %		
Yes	5.9	—
No	94.1	—
Sleep deprivation level, mean ± SD (1 = none, 5 = extreme)	2.5 ± 0.70	—
Stress level, mean ± SD (1 = none, 5 = extreme)	2.76 ± 0.74	—

reliable performance assessment tool. Similar findings have been demonstrated in other studies where performance checklists were used.[21] Second, tools developed for aviation may not be transferable to medical domains. As Flin and Maran[17] have pointed out, "it is not sufficient to take aviation training materials and simply delete 'pilot' and replace it with 'nurse' or 'anesthetist.'" Third, it may not be appropriate to adapt evaluation tools from one medical context to another without taking into account key differences that may exist. Obstetric crisis management may be sufficiently unique to require a domain specific evaluation tool.

Self-assessed team scores of the HFRS by participants were not able to reliably assess team performances in that they were unable to discriminate between good and bad performances. Nor were the HFRS self-assessed scores able to identify difficulty of the scenario itself. In addition, discrepancy between rater groups was noticed in that the nurses tended to be more generous with their team self-assessments than physician raters. Similarly, HFRS scores generated by external observers viewing videotapes of the performances required a large number of ratings (nine independent raters) to produce values with reliabilities in the 0.80 range, the range usually accepted as necessary for effective discrimination of performance. Therefore, if one were to try to evaluate team performance on a larger scale, self-ratings using the HFRS would be problematic, and the number of external raters needed to generate reliable scores on the HFRS might very well be prohibitive.

In contrast, the GRS, whether produced by external examiners or self-ratings, was better able to differentiate team performances, was better able to distinguish between scenarios of differing difficulty, and did not demonstrate differences between raters' self-assessments as a function of the rater's profession. Self-generated global ratings were moderately reliable when averaging the six team members' scores (0.76), and when using external examiners watching videotapes, even this simple global scale could achieve reliabilities of 0.8 with as few as six independent examiners. Reasons for these findings include the fact that raters had only one score with which to agree or disagree, and the GRS provided a more consistent method to rate performance in that raters were able to consider the outcome of the exercise as a measure of team performance. In fact, the rating scales may not be measuring similar competence domains. Global rating scales have been shown to be useful assessments of individual performances and could potentially be more useful than checklists for evaluation purposes.[22-25] The limitation of the global rating scale used for this study lies in its simplicity. To provide valuable feedback, the GRS would not identify areas for team training, unlike a checklist that would allow such definition. Because the debriefing of the team performance is crucial to the learning and safety outcomes, the GRSs

would have to be adapted to provide more information on curricular areas to be addressed. Therefore, although potentially useful as a summative evaluation tool, the GRS in this study has limitations in use as a formative tool. If GRSs are to be used as formative evaluation tools, they would have to be expanded to include a few key subcategories that would allow assessors to provide more specific feedback to participants during the debriefing.

The moderate to good reliability of the GRS does raise questions about the lack of reliability in the HFRS. That is, the reliability of the GRS suggests that there are measurable differences between the teams' performances that can be captured by a fairly simple rating scale, so the fact that the HFRS was unable to do so suggests a problem with the scale itself.

Although further adaptation of the HFRS may prove to be reliable, there is sufficient evidence from the results of this study to warrant the development of an HFRS from first principles for assessment of obstetric team performance. Using qualitative analysis of safety attitudes from focus groups as well as expert and nonexpert opinions about behavioral markers demonstrated during the obstetric team management in a simulated environment, lists of behaviors can be generated and categorized for use as human factors performance items. Similar methodology has been used to develop behavioral marking systems for both anesthesiologists and surgeons.[3,14] In addition, review of the literature may reveal marking systems that have been used, and these can be examined for common themes for inclusion in a newly developed marking system.[5,26] When a marking system has been developed, it can then be pilot tested on the specific groups to which it will apply in order to assess validity and reliability of the tool.

We chose to use both self-assessments and externally generated assessments of team performance. The ability to critically examine one's own strengths, especially within the context of a team, has been touted as a powerful tool for self-directed learning.[27] There have been some studies that suggest effective self-assessment ability in professionals. For example, self-assessment in simulation-based surgical skills training of novice learners has indicated that self-assessments reflect actual performance.[28] Similarly, a few studies of postgraduate trainees and expert surgeons indicate the reliability of self-assessment to observed performance.[29,30] In contrast, our findings showed a relatively small correlation between the self-assessment and externally generated assessments of performance when using either the HFRS or the GRS. This fairly small correlation is, in fact, consistent with an extensive body of literature that questions the use of self-assessment as a valid measure of actual performance.[31-34] Our data, therefore, reinforce the need to use external raters regardless of the measurement instrument being used to assess performance.

Although the investigators could not completely simulate all of the normal occurrences that arise during the development of an urgent or emergent event in the delivery room, *e.g.*, movement of the patient from a labor room to the operating room, the scenarios did involve a "handing" over process or a situation in which the attending nurse was required to summon the team and communicate the sequence of events to others who arrived. Therefore, issues such as leadership and communication, often established during transfer, were still incorporated into the scenario before any specific surgical or anesthetic intervention occurred.

One of the strengths of our study design is the introduction of an obstetric model allowing for a more realistic participation of the surgeons in the simulated scenario. This is the first published report in the literature of a high-fidelity simulation of obstetric team performance with anesthesiologists, nurses, and obstetricians involved in the hands-on management of obstetric crises. Traditionally, the anesthesiologists' simulated work environment has been shown to be a high-fidelity representation, but actors provided the roles of surgeons, nurses, and other medical personnel. This study allowed for genuine interaction between participants from different disciplines and professions.

The findings of this study identify a need for the development of a domain specific behavioral marking tool for obstetric teams. It is our intention to use the findings from this study to develop such a behavioral marking system from first principles and to address the issues of validity and reliability of the newly developed tool.

References

1. Kohn LT, Corrigan JM, Donaldson MS: To Err Is Human: Building a Safer Health System. Edited by Kohn LT, Corrigan JM, Donaldson MS. Washington, DC, National Academy Press, 1999, pp 1–67

2. Helmreich R, Schaefer H-G: Team performance in the operating room, Human Error in Medicine. Edited by Bognor M. Mahwah, New Jersey, Lawrence Erlbaum Associates, 1994, pp 225–53

3. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R: Rating non-technical skills: Developing a behavioural marker system for use in anaesthesia. Cogn Tech Work 2004; 6:165–71

4. Weiner E, Kanki B, Helmreich R: Cockpit Resource Management. San Diego, Academic Press, 1993

5. Reader T, Flin R, Lauche K, Cuthbertson B: Non-technical skills in the intensive care unit. Br J Anaesth 2006; 96:551–9

6. Nunnally J, Bernstein I: Psychometric Theory, 3rd edition. New York, McGraw Hill, 1993

7. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R: Anaesthetists' non-technical skills (ANTS): Evaluation of a behavioural marker system. Br J Anaesth 2003; 90:580–8

8. Yule S, Flin R, Paterson-Brown S, Maran N, Rowley D: Development of a rating system for surgeon's non-technical skills. Med Educ 2006; 40:1098–104

9. Schaefer H-G, Helmreich R: The Operating Room Management Attitudes' Questionnaire (ORMAQ). Austin, National Aeronautics and Space Administration/University of Texas, 1993

10. Sexton J, Helmreich RL, Glenn D, Wilhelm J, Merritt A: Operating Room Management Attitudes Questionnaire (ORMAQ). Austin, Human Factors Research Project, University of Texas at Austin, 2000

11. Helmreich R, Merritt A, Sherman P: The Flight Management Attitudes Questionnaire (FMAQ). Austin, National Aeronautics and Space Administration, 1993, pp 93–4

12. Helmreich R, Davies J: Human factors in the operating room: Interpersonal determinants of safety, efficiency and morale. Baillieres Clin Anaesthesiol 1996; 10:277–95

13. Flin R, Fletcher G, McGeorge P, Sutherland A, Patey R: Anaesthetists' attitudes to teamwork and safety. Anaesthesia 2003; 58:233–42

14. Flin R, Yule S, McKenzie L, Paterson-Brown S, Maran N: Attitudes to teamwork and safety in the operating theatre. Surgeon 2006; 4:145–51

15. Sexton B, Helmreich R, Neilands T, Rowan K, Vella K, Boyden J, Roberts P, Thomas E: The Safety Attitudes Questionnaire: Psychometric properties, benchmarking data and emerging research. BMC Health Serv Res 2006; 6:44–54

16. Cooper G, McClure J: Why Mothers Die 2000-2002: Executive Summary and Key Findings. London, Royal College of Obstetricians and Gynaecologists, 2005, pp 3–15

17. Flin R, Maran N: Identifying and training non-technical skills for teams in acute medicine. Qual Saf Health Care 2004; 13 (suppl):i80–4

18. Howard SK, Gaba DM, Fish KJ, Yang G, Sarnquist FH: Anesthesia crisis resource management training: Teaching anesthesiologists to handle critical events. Aviat Space Environ Med 1992; 63:763–70

19. Gaba D, Howard S, Flanagan B, Smith B, Fish K, Botney R: Assessment of clinical performance during simulated crises using both technical and behavioral ratings. ANESTHESIOLOGY 1998; 89:8–18

20. Healey A, Undre S, Vincent C: Developing observational measures of performance in surgical teams. Qual Saf Health Care 2004; 13 (suppl):i33–40

21. Hodges B, Regehr G, McNaughton N, Tiberius R, Hanson M: OSCE checklists do not capture increasing levels of expertise. Acad Med 1999; 74:1129–34

22. Morgan PJ, Cleave-Hogg D, Guest CB: A comparison of global ratings and checklist scores in undergraduate anesthesia simulator assessments. Acad Med 2001; 76:1053–5

23. Regehr G, MacRae H, Reznick RK, Szalay D: Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med 1998; 73:993–7

24. Regehr G, Freeman R, Hodges B, Russell L: Assessing the generalizability of OSCE measures across content domains. Acad Med 1999; 74:1320–2

25. Vassiliou M, Feldman L, Andrew C, Bergman S, Leffondre K, Stanbridge D, Fried G: A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005; 190:107–13

26. Yule S, Flin R, Paterson-Brown S, Maran R: Non-technical skills for surgeons in the operating room: A review of the literature. Surgery 2006; 139:140–9

27. Regehr G, Hodges B, Tiberius R, Lofchy J: Self-education for professionals. Acad Med 1996; 71:S52–4

28. MacDonald J, Williams R, Rogers D: Self-assessment in simulation-based surgical skills training. Am J Surg 2003; 185:319–22

29. Mandel L, Goff B, Lentz G: Self-assessment of resident surgical skills: Is it feasible? Am J Obstet Gynecol 2005; 193:1817–22

30. Sarker S, Hutchinson R, Chang A, Vincent C, Darzi A: Self-appraisal hierarchical task analysis of laparoscopic surgery performed by expert surgeons. Surg Endosc 2006; 20:636–40

31. Falchikov N, Boud D: Student self-assessment in higher education: A meta-analysis. Rev Educ Res 1989; 59:395–430

32. Gordon M: A review of the validity and accuracy of self-assessments in health professions training. Acad Med 1991; 66:762–9

33. Ward M, Gruppen L, Regehr G: Research in self-assessment: Current state of the art. Adv Health Sci Educ 2002; 7:63–80

34. Dunning D, Heath C, Suls J: Flawed self-assessment: Implications for health, education and the workplace. Psychol Sci Public Interest 2004; 5:69–106

Appendix A: Scenarios

Scenario 1

Morbidly obese parturient with nonreassuring fetal heart rate trace; decision to perform cesarean section: difficult intubation, hypoxemia, and cardiac arrest

History and Physical

This 32-yr-old *gravida 1 para 0* parturient at 37 weeks gestation arrived on the labor floor complaining of regular painful contractions. The membranes are intact and she is Group B streptococcus positive. She has had an uneventful pregnancy to date and review of her past health is unremarkable. She has no known drug allergies and takes no medications.

On examination, she is a morbidly obese woman with a height of 160 cm and weight 130 kg. Her blood pressure is 130/80 mmHg and heart rate 120 beats/min. She is afebrile. She is in obvious pain. On admission, pelvic exam reveals a cervical dilation of 1 cm. The monitor shows every 2–3 min contractions lasting 45 s. The fetal heart rate tracing shows variable decelerations. After 4 h, repeat cervical exam reveals a cervical dilation of 1 cm and progressively severe decelerations.

No epidural in place as patient adamantly refused regional anesthesia

Backup nurse in the operating room with the shift nurse

Backup nurse explains to the shift nurse that she was told by the obstetric staff to take the patient to the operating room for a possible cesarean section due to fetal decelerations

Backup nurse leaves

Fetal decelerations continue

Nurse calls for obstetrician and resident

Nurse calls for a backup nurse

Anesthesia staff and resident are called

Backup nurse arrives

Anesthesia Arrives

Obstetrician arrives

Continuing audible fetal decelerations

A lengthy deceleration makes it necessary for anesthesia to start a general anesthetic

Induction started

Anesthesia attempts to intubate, but airway is impossible to intubate using laryngoscopy (chart lists the airway exam as a Mallampati II)

Attempts at bag–mask ventilation will fail

Anesthesia calls for the fiberoptic bronchoscope

Fiberoptic arrives with no light source

Backup nurse must search for the light source

(If anesthesia attempts a laryngeal mask airway; adequate ventilation will not be possible)

Continuing severe fetal decelerations

Patient becoming increasingly hypoxic

Patient develops pulseless electrical activity

Scenario 2

Urgent cesarean section for parturient with worsening preeclampsia: critical event involving severe hypertension and pulmonary edema

Epidural already in place

Patient is already in the operating room with a nurse (patient kind of difficult and annoying)

Husband also in the room (very annoying)

Nurse calls obstetrician to inform them that the patient is in the operating room

Anesthesia and resident called

Obstetrician and anesthesia arrive

Obstetrician tests epidural and finds a "patchy" block

Anesthesia attempts to top up the epidural for the cesarean section

Patient's blood pressure remains high (220/115)

Epidural eventually works, but blood pressure remains high

Blood pressure resistant to drugs

Delay in getting started due to the patchy block

Cesarean section starts

Obstetrician having difficulty extracting baby, asks for nitroglycerin

Husband and patient continue to be very annoying

Patient starts to desaturate, complains of dyspnea, restless, becoming more hypertensive

Patient developing pulmonary edema

Husband becoming increasing worried and refuses to leave

Scenario 3

Emergency cesarean section in parturient with twin gestation at 34 weeks for cord prolapse: critical event amniotic fluid embolism

Patient in operating room for stat cesarean section

Nurse in room with patient

No epidural in place—patient refused

Anaesthesia and obstetrician called

Backup nurse called

Anaesthesia staff arrives

Induction of general anesthesia started quickly for cord prolapse

As soon as obstetrician arrives, he or she continually gets paged to help out in another operating room for massive blood loss

Cesarean section starts

Obstetrician and resident deliver babies

As soon as babies out, blood pressure drops, CO_2 drops, Sp_{O_2} drops

Patient develops asystole

Scenario 4

Profound fetal bradycardia secondary to occult abruptio placenta: critical event involving massive blood loss and inability to maintain blood pressure

Patient in operating room due to vaginal bleeding

Nurse in room with patient

Profound fetal bradycardia

Anaesthesia and obstetrician called (staff and resident)

Anaesthesia resident arrives first and starts the general anesthetic (must start right away due to profound fetal bradycardia)

Obstetrician arrives (obstetric resident doesn't arrive, in another cesarean section)

General anesthetic started

Preinduction vitals: blood pressure 110/50, heart rate 130 beats/min

Postinduction vitals: blood pressure 90/60, heart rate 140 beats/min

Cesarean section begins

Massive blood loss

Blood pressure 60 systolic

Obstetrician calls for backup but will take 20–30 min to arrive

Nurse calls blood bank for cross and type 4 units of blood

Ongoing blood loss

Blood not coming; nurse calls blood bank and discovers that the porter has left with the blood but has not arrived

Appendix B: Human Factors Rating Scale

With respect to the team performance you are witnessing, please complete the survey using the following scale:

1 = strongly disagree, 2 = slightly disagree, 3 = no opinion, 4 = slightly agree, 5 = strongly agree

If the question does not apply to the scenario, please leave blank.

Leadership	**1**	**2**	**3**	**4**	**5**

1. The obstetrician encouraged questions from the obstetric resident.
2. The anesthesiologist encouraged questions from the anesthesia resident.
3. The successful management of the scenario was mainly a function of the obstetrician's expertise.
4. The successful management of the scenario was mainly a function of the anesthesiologist's expertise.
5. The obstetric resident should have been more involved in the patient's care.
6. The anesthesia resident should have been more involved in the patient's care.
7. The successful management of the case was mainly due to the technical proficiency of the physicians.
8. During the critical event management, the nurses were appropriately consulted by the physicians.
9. The nurses assumed a leadership role during the scenario.

Confidence–Assertion	**1**	**2**	**3**	**4**	**5**

10. During the critical event management, the most senior physician was in charge.
11. The residents questioned the actions of the consultant physicians.
12. When a critical event occurred, the nurses(s) questioned the actions of the physician.
13. When a critical event occurred, the obstetrician questioned the actions of the anesthesiologist.
14. When a critical event occurred, the anesthesiologist questioned the actions of the obstetrician.
15. If there was uncertainty on the part of any team member, the team openly questioned that team member on his or her actions.

Information Sharing	**1**	**2**	**3**	**4**	**5**

16. The attending anesthesiologist clearly verbalized his/her plan of action when a critical event occurred.
17. The attending obstetrician clearly verbalized his or her plan of action when a critical event occurred.
18. The anesthesiologist ensured that requests for action were acknowledged by the receiver.
19. The obstetrician ensured that requests for action were acknowledged by the receiver.
20. The nurse(s) ensured that requests for action were acknowledged by the receiver.
21. The obstetric resident ensured that requests for action were acknowledged by the receiver.
22. The anesthesia resident ensured that requests for action were acknowledged by the receiver.
23. The team shared information efficiently.
24. Communication between physicians was effective.
25. Communication between physicians and nurses was effective.
26. Communication between nurses was effective.

Teamwork	**1**	**2**	**3**	**4**	**5**

27. Obstetricians gave feedback to the anesthesiologist.
28. Anesthesiologists gave feedback to the obstetricians.
29. Physicians gave feedback to the nurses.
30. Nurses gave feedback to the physicians.
31. The anesthesiologist took charge of coordinating the team effort.
32. The obstetrician took charge of coordinating the team effort.
33. The nurses took charge of coordinating the team effort.
34. The team effectively prioritized activities.
35. Conflicts were openly resolved.
36. The team worked well together.

Error	**1**	**2**	**3**	**4**	**5**

37. Basic rules were broken during the management of the case.
38. Mistakes were made and not voiced by team members.
39. Errors that were committed resulted from lack of knowledge.
40. Errors that were committed resulted from lack of communication.
41. Errors that were committed resulted from lack of equipment.
42. Errors that were committed resulted from lack of technical skills.
43. Errors that were committed resulted from lack of following guidelines.
44. Errors that were committed resulted from lack of experience.
45. Errors that were committed resulted from lack of resources.

Appendix C: Global Rating of Team Performance

☐ ☐ ☐ ☐ ☐
1 2 3 4 5
Unacceptable Borderline Acceptable Good Superior
Performance Performance Performance Performance Performance

Categories and Descriptors

- 1 = Unacceptable Performance
- Multiple errors which may have or did lead to irreversible damage to the patient
- Did not recognize more than one critical event without assistance
- A large number of unplanned errors committed
- No team communication

- 2 = Borderline Performance
- Many errors which had the potential to lead to irreversible damage to the patient but were recognized by the team and corrected
- Slow response to critical events with some assistance required
- A few unplanned errors committed
- Poor team communication

- 3 = Acceptable Performance
- A number of errors that would not have led to irreversible damage
- Recognized all the critical events but relatively slow response time to recognition and treatment
- A few unplanned errors committed
- Satisfactory team communication but lacking in leadership

- 4 = Good Performance
- A few errors that were minor in nature and did not pose a serious risk to the patient
- Recognized critical events and responded in an acceptable time frame
- A few unplanned errors that were corrected
- Good team communication

- 5 = Superior Performance
- Very few errors that were minor in nature and did not pose a serious risk to the patient
- Prompt recognition and management of critical events
- No unplanned errors committed
- Excellent leadership with clear, concise team communication